DOCUMENT RESUME

ED 167 634 TM 008 474

AUTHOR Parlett, Malcolm; Hamilton, David
TITLE "Evaluation as Illumination: A New Approach to the Study of Innovatory Programs". Occasional Paper.
INSTITUTION Edinburgh Univ. (Scotland). Centre for Research in the Educational Sciences.
SPONS AGENCY Nuffield Foundation, London (England).
PUB DATE Oct 72
NOTE 35p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Educational Alternatives; Educational Anthropology; *Educational Environment; *Educational Innovation; *Evaluation Methods; Foreign Countries; Formative Evaluation; Observation; *Program Evaluation; *Research Methodology; Research Problems; Scientific Methodology
IDENTIFIERS Great Britain; Illuminative Evaluation

ABSTRACT

Conventional approaches to program evaluation have followed the experimental and psychometric traditions dominant in educational research. Their aim (unfulfilled) of achieving fully objective methods had led to studies that are artificial and restricted in scope. Illuminative evaluation is introduced as a total re-appraisal of the rationale and techniques of program evaluation. It follows a contrasting 'anthropological' research paradigm. Illuminative evaluation recognizes that an innovative instructional program cannot be separated from its learning milieu--a network of cultural, social, institutional, and psychological forces. The evaluator concentrates on processes within the classroom rather than on outcomes (goals, objectives) derived from a specification of the instructional system. Data is collected from observation, interview, questionnaires, and documentary information. Illuminative evaluation thus concentrates on the information gathering rather than on the decision making component of evaluation. The paper concludes with a discussion of the problems and potentialities of the new approach: its range of applicability; objectivity; validity; obligations of the research worker; how explorations of the learning milieu can meet the need for theoretical advance; and how the illuminative approach can clarify decision-making. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

- ED167634

Centre for Research in the Educational Sciences

University of Edinburgh.

OCCASIONAL PAPER

October, 1972

"EVALUATION AS ILLUMINATION: A NEW APPROACH TO THE STUDY OF INNOVATORY PROGRAMS"

by

MALCOLM PARLETT AND DAVID HAMILTON

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Malcolm Parlett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM008 474

© Malcolm Parlett and David Hamilton, 1972.

Innovation is now a major educational priority. For nearly two decades it has expanded and proliferated. It absorbs increasing sums of public and private money. Its impact is felt throughout the world.[1] Curricula are restructured, new devices introduced, forms of teaching permutated. But decisions to change are more than educational: questions of politics, ideology, fashion, and finance also intervene.

More recently - to aid decision-making - innovation has been joined by evaluation. Increasingly, committees and foundations fund evaluation studies as an integral part of innovation programs. Like innovation itself, evaluation has rapidly developed a legitimacy and importance of its own: professional journals have been launched and research centers established. The 'evaluator' has emerged as a new and influential figure. In short, both innovation and evaluation have become 'big science'.[2]

As a new field, program evaluation has encountered a wide range of problems, both theoretical and methodological. Current concerns[3] include the 'roles' of evaluation;[4] the neutrality of the

---

1. See for example the series of reports: Innovation in Education (Paris: O.E.C.D. Centre for Educational Research and Innovation, 1971), which reviews developments in member countries.

2. D.J. de S. Price, Little Science, Big Science (Columbia: Columbia University Press, 1963).

3. Much of this debate has been monitored by the AERA Monograph Series on Curriculum Evaluation (Chicago: Rand McNally, 1967 onwards).

4. M. Scriven, "The methodology of evaluation," in R. Tyler et al., Perspectives of Curriculum Evaluation, AERA Monograph No. 1, 1967; pp. 39-83.

evaluator;[5] the value of classroom observation;[6] the function of 'formative' evaluation;[7] the use of 'objectives';[8] and the value of long-term studies.[9] Confusion is engendered as rival proposals, models and terminologies are voiced and then rapidly countered. As a developing field of study, evaluation proceeds in the absence of coherent or agreed frames of reference.

More generally within educational research two distinct paradigms[10] can be discerned. Each has its own strategies, foci and assumptions. Dominant is the 'classical' or 'agricultural-botany'[11] paradigm, which

---

5. F.G. Caro, "Issues in the evaluation of social programs," Review of Educational Research, 41, 1971; pp. 87-114.

6. R.J. Light & P.V. Smith, "Choosing a future: strategies for designing and evaluation new programs," Harvard Educational Review, 40, Winter, 1970; pp. 1-28.

7. L.M. Smith, "Participant observation and evaluation strategies," paper presented to AERA symposium on "Participant Observation and Curriculum: Research and Evaluation" (New York, February 1971).

8. J. Popham et al., Instructional Objectives, AERA Monograph No. 3, 1963.

9. F.G. Caro, op. cit.

10. The term paradigm as used by T.S. Kuhn, (The Structure of Scientific Revolutions, Chicago: University of Chicago Press, (2nd ed.) 1970) is an overarching concept similar in meaning to 'world-view', 'philosophy', or even 'intellectual orthodoxy'. A paradigm prescribes problem fields, research methods, and acceptable standards of solution and explanation for the academic community it embraces.

11. M. Parlett, "Evaluating innovations in teaching," in H.J. Butcher and E. Rudd (eds.), Contemporary Problems in Higher Education (London: McGraw-Hill, 1972). The designation 'agricultural-botany' is not fortuitous. Many of the statistical and experimental techniques used in educational research were originally developed (e.g. by Fisher) for use in agricultural experimentation.

utilizes a hypothetico-deductive methodology derived from the experimental and mental-testing traditions in psychology. Almost all evaluation studies have resided within this traditional paradigm.

More recently, a small number of empirical studies have been conceived outside the agricultural-botany framework, and relate instead to social anthropology, psychiatry, and participant observation research in sociology.[12] Such research can be thought of as representing a second and contrasting paradigm, with a fundamentally different research style and methodology from that of mainstream educational research. We outline here an approach to evaluation that belongs to this alternative, or 'social anthropology' paradigm.[13]

TRADITIONAL EVALUATION AND THE AGRICULTURAL-BOTANY PARADIGM

The most common form of agricultural-botany type evaluation is presented as an assessment of the effectiveness of an innovation by examining whether or not it has reached required standards on

---

12. See for example, J. Henry, Essays on Education (Harmondsworth: Penguin Books, 1971); P.W. Jackson, Life in Classrooms (New York: Holt, Rinehart & Winston, 1968) and M.F.D. Young (ed.), Knowledge and Control (London: Collier-Macmillan, 1971).

13. M. Parlett, op. cit.

pre-specified criteria.[14] Students- rather like plant crops - are given pre-tests (the seedlings are weighed or measured) and then submitted to different experiences (treatment conditions). Subsequently, after a period of time, their attainment (growth or yield) is measured to indicate the relative efficiency of the methods (fertilizers) used. Studies of this kind are designed to yield data of one particular type, i.e. 'objective' numerical data that permit statistical analyses.[15] Isolated variables like I.Q., social class, test scores, personality profiles and attitude ratings are codified and processed to indicate the efficiency of new curricula, media or methods.

---

14. For instance within this framework Lindvall and Cox have argued that the 'effectiveness' of an innovation is 'determined' by the answers to four basic questions. '(1) What goals should the program achieve? (2) What is the plan for achieving these goals? (3) Does the operating program represent a true implementation of the plan? (4) Does the program, when developed and put into operation, achieve the desired goals?' (C.M. Lindvall & R.C. Cox, The IPI Evaluation Program, AERA Monograph, No. 5, 1970; pp. 5-6.) At face value these questions seem reasonable. But they embody problematic assumptions. For example, programs rarely have clearly specified and commonly agreed 'desired goals'. Measurement of 'goal achievement' is never unequivocal. To speak of a 'true implementation' is utopian, even nonsensical in terms of educational practice.

15. Again, within this framework, it has been proposed that 'the search for (compensatory) programs which are working well should become a three-step procedure: (1) First, locate the best recreatable centers, employing techniques which use estimates of average random variation. This will require analysis of variance, followed by multiple comparison procedures. (2) Estimate, separately, the impact of random factors upon only the best-performing centers. (3) Use this selective estimate of random variation to test whether these best centers are outperforming chance, and are worth recreating.' (R.J. Light & P.V. Smith, op. cit. p. 18.) The evaluation suggested by Light and Smith is based purely on numerical results; no other data would be collected or eligible for consideration.

Recently, however, there has been increasing resistance to evaluations of this type.[16] The more notable shortcomings may be briefly summarized as follows:

1. Educational situations are characterized by numerous relevant parameters. Within the terms of the agricultural-botany paradigm these must be randomized using very large samples; or otherwise strictly controlled. The former approach entails a major data collection exercise and is expensive in time and resources. It also runs counter to the need, widely acknowledged, for evaluation before large scale application rather than after it. The latter procedure - of strict control - is rarely followed. To attempt to simulate laboratory conditions by 'manipulating educational personnel' is not only dubious ethically, but also leads to gross administrative and personal inconvenience. Even if a situation could be so unnervingly controlled, its artificiality would render the exercise irrelevant: rarely can 'tidy' results be generalized to an 'untidy' reality. Whichever approach is used, there is a tendency for the investigator

---

16. The objections are developed more extensively in: M. Guttentag, "Models and methods in evaluation research," J. Theory Soc. Behavior, 1, 1971; pp. 75-95: R.E. Stake, "Measuring what learners learn (with a special look at performance contracting)," (University of Illinois at Urbana-Champaign: Center for Instructional Research and Curriculum Evaluation, n.d.); B. MacDonald, "The evaluation of the Humanities Curriculum Project: a wholistic approach," (University of East Anglia, Norwich: Center for Applied Research in Education, n.d.); S. Messick, "Evaluation of educational programs as research on educational process," (Princeton: Educational Testing Service, n.d.); L.C. Taylor, Resources for Learning (Harmondsworth: Penguin Books, 1971); and M. Parlett, op. cit..

to think in terms of 'parameters' and 'factors' rather than 'individuals' and 'institutions'. Again, this divorces the study from the real world.

2. Before-and-after research designs assume that innovatory programs undergo little or no change during the period of study. This built-in premise is rarely upheld in practice. Yet it remains fundamental to the design, constraining the researchers from adapting to the changed circumstances that so frequently arise.[17] It may even have a deleterious effect on the program itself, by discouraging new developments and re-definitions mid-stream. Longitudinal studies, for these reasons, rarely can serve an effective 'formative' or cybernetic function.[18]

3. The methods used in traditional evaluations impose artificial and arbitrary restrictions on the scope of the study. For instance,

---

17. During the long-term evaluation of a Swedish individualized math program (IMU), teachers who wished to transfer students from one class to another were actively discouraged by the evaluator from doing so, on the grounds that it would render inter-class comparisons invalid. Teachers also requested separate diagnostic tests for less able children. But again, this educational need was subordinated to the evaluator's requirements. The British evaluation of the initial teaching alphabet (i.t.a.), set up to compare pupil progress with i.t.a. and traditional orthography (t.o.) over a five year period, provides a second example. Early before-and-after test results indicated that a major educational difficulty was faced by children transferring from i.t.a. to t.o.. But nothing was done to focus attention on this problem: the research die had already been cast. (See J. Downing (ed.), The i.t.a. Symposium, London: National Foundation for Educational Research, 1967).

18. Because the pre-specification of parameters, by definition, occurs at the outset, variables which emerge during the study are likely to be left out of the analysis. In an extreme case this neglect of 'new' variables may negate an entire evaluation study. After criticism that it had not controlled for the Hawthorne effect, the i.t.a. experiment was re-started after two years.

the concentration on seeking quantitative information by objective means can lead to neglect of other data, perhaps more salient to the innovation, but which is disregarded as being 'subjective', 'anecdotal', or 'impressionistic'. However, the evaluator is likely to be forced to utilize information of this sort if he is satisfactorily to explain his findings, weight their importance, and place them in context.

4. Research of this type, by employing large samples and seeking statistical generalizations, tends to be insensitive to local perturbations and unusual effects. Atypical results are seldom studied in detail. Despite their significance for the innovation, or possible importance to the individuals and institutions concerned, they are ironed out and lost to discussion.

5. Finally, this type of evaluation often fails to articulate with the varied concerns and questions of participants, sponsors, and other interested parties. Since classical evaluators believe in an 'objective truth' equally relevant to all parties, their studies rarely acknowledge the diversity of questions posed by different interest-groups.

These points suggest that applying the agricultural-botany paradigm to the study of innovations is often a cumbersome and inadequate procedure.[19] The evaluation falls short of its own tacit claims to be

---

19. We are not, of course, arguing here against the use of experimental longitudinal or survey research methods as such. Rather, for the reasons suggested, we submit that they are usually inappropriate, ineffective, or insufficient for program evaluation purposes.

controlled, exact and unambiguous. Rarely, if ever, can educational programs be subject to strict enough control to meet the design's requirements. Innovations, in particular, are vulnerable to manifold extraneous influences. Yet the traditional evaluator ignores these. He is restrained by the dictates of his paradigm to seek generalized findings along pre-ordained lines. His definition of empirical reality is narrow. One effect of this is that it diverts attention away from questions of educational practice towards more centralized bureaucratic concerns.

ILLUMINATIVE EVALUATION AND THE SOCIAL-ANTHROPOLOGY PARADIGM

Although traditional forms of evaluation have been criticized in this way, little attempt has been made to develop alternative models. The model described here, illuminative evaluation,[20] takes account of the wider contexts in which educational programs function. Its primary concern is with description and interpretation rather than

---

20. The term illuminative research is drawn from M.A. Trow, "Methodological problems in the evaluation of innovation," in M.C. Wittrock & D.E. Wiley (eds.), The Evaluation of Instruction (New York: Holt, Rinehart & Winston, 1970; pp. 289-305). The approach to evaluation described here grew out of research at MIT in association with B.R. Snyder and M.J. Kahne. (See M. Parlett, "Undergraduate teaching observed," Nature, 223, 1969; pp. 1102-1104.)

measurement and prediction.[21] It stands unambiguously within the alternative anthropological paradigm. The aims of illuminative evaluation are to study the innovatory program: how it operates; how it is influenced by the various school situations in which it is applied; what those directly concerned regard as its advantages and disadvantages; and how students' intellectual tasks and academic experiences are most affected. It aims to discover and document what it is like to be participating in the scheme, whether as teacher or pupil; and, in addition, to discern and discuss the innovation's most significant features, recurring concomitants, and critical processes. In short, it seeks to address and to illuminate a complex array of questions: 'Research on innovation can be enlightening to the innovator and to the whole academic community by clarifying the processes of education and by helping the innovator and other interested parties to identify those procedures, those elements in the educational effort, which seem to have had desirable results.'[22]

The paradigm shift entailed in adopting illuminative evaluation requires more than an exchange of methodologies: it also involves new suppositions, concepts, and terminology. Central to an understanding

---

21. For three published reports which approach the style advocated here see: J.P. Hanley et al., Curiosity, Competence, Community (Cambridge, Mass.: Educational Development Center Inc., 1969). L.M. Smith & P.A. Pohland, "Technology and the rural highlands," AERA Monograph No. 8, 1971. And M. Parlett & J.G. King, Concentrated Study, Research in Higher Education Monograph No. 14 (London: Society for Research in Higher Education, 1971).

22. M.A. Trow, op. cit. p. 302.

of illuminative evaluation are two concepts: the 'instructional system' and the 'learning milieu'.

The Instructional System

Educational catalogs, prospectuses, and reports charactistically contain a variety of formalized plans and statements which relate to particular teaching arrangements. Each of these summaries can be said to constitute or define an instructional system; and includes, say, a set of pedagogic assumptions, a new syllabus, and details of techniques and equipment. This 'catalog description' is an idealized specification of the scheme: a set of elements arranged to a coherent plan. Despite their immense variation, the Dalton Plan, performance contracting, programmed learning, the integrated day, team teaching, 'Sesame Street' and 'Man: A Course of Study' can all be considered as instructional systems in these terms.

The traditional evaluator builds his study around innovations defined in this way. He examines the blueprint or formalized plan and extracts the programs' goals, objectives, or desired outcomes. From these, in turn, he derives the tests and attitude inventories he will administer. His aim is to evaluate the instructional system by examining whether, for example, it has 'attained its objectives' or met its 'performance criteria'.

This technological approach fails to recognize the catalog description for what it is. It ignores the fact that an instructional system, when adopted, undergoes modifications that are rarely trivial. The instructional system may remain as a shared idea, abstract model,

slogan, or shorthand, but it assumes a different form in every situation. Its constituent elements are emphasised or de-emphasised, expanded or truncated, as teachers, administrators, technicians, and students interpret and re-interpret the instructional system for their particular setting. In practice, objectives are commonly re-ordered, re-defined, abandoned or forgotten. The original 'ideal' formulation ceases to be accurate, or indeed of much relevance. Few in practice take catalog descriptions and lists of objectives very seriously, save - it seems - for the traditional evaluator.

To switch from discussing the instructional system in abstract form to describing the details of its implementation is to cross into another realm. Here the second new concept is required.

The Learning Milieu

This is the social-psychological and material environment in which students and teachers work together. The learning milieu represents a network or nexus of cultural, social, institutional, and psychological variables. These interact in complicated ways to produce, in each class or course, a unique pattern of circumstances, pressures, customs, opinions, and work styles which suffuse the teaching and learning that occur there. The configuration of the learning milieu, in any particular classroom, depends on the interplay of numerous different factors. For instance, there are numerous constraints (legal, administrative, occupational, architectural and financial) on the organization of teaching in schools; there are pervasive operating assumptions (about the arrangement of subjects, curricula, teaching

methods, and student evaluation) held by faculty: there are the individual teacher's characteristics (teaching style, experience, professional orientation, and private goals); and there are student perspectives and preoccupations.

Acknowledging the diversity and complexity of learning milieux is an essential pre-requisite for the serious study of educational programs. The argument advanced here is that innovatory programs, even for research purposes, cannot sensibly be separated from the learning milieux of which they become part. If an evaluation study hinges on the supposed perpetuation of the instructional system in more or less its original form, it makes an arbitrary and artificial distinction: it treats the innovation as a self-contained and independent system, which in practice it is manifestly not.

The introduction of an innovation sets off a chain of repercussions throughout the learning milieu. In turn these unintended consequences are likely to affect the innovation itself, changing its form and moderating its impact. For example, at the Massachusetts Institute of Technology, it was found that switching from 'distributed' to 'concentrated' study (a change from students taking several subjects concurrently to intensive full-time study of a single subject) was, in the event, far more than a re-scheduling arrangement.[23] It demanded new pedagogic forms (continuous lecturing would have led to 'overload'); it resulted in new role relationships between faculty and students (daily

---

23. M. Parlett & J.G. King, op. cit.

contact encouraged a degree of informality impossible with two meetings a week of one hour each); and it changed peer relations between students (their working alongside the same students continuously led to much greater interaction than is usual in M.I.T. sophomore classes). Such profound shifts in the learning milieu produced a further range of important secondary effects, apparently far removed from the innovation as such, but ultimately deriving from it.

To attempt to gauge the impact of the innovation (in this instance 'concentrated study') without paying attention to factors such as these, would clearly be absurd. In the above study it was possible to trace how each of these milieu effects had its corollary in the intellectual sphere: e.g. the informality encouraged normally silent students to ask questions; and though the range of different learning activities was regarded as excellent for achieving basic comprehension of the subject-matter, it might have put the students at a disadvantage in a conventional exam.

Connecting changes in the learning milieu with intellectual experiences of students is one of the chief concerns for illuminative evaluation. Students do not confront 'knowledge' in naked form; it comes to them clothed in texts, lectures, tape-loops, etc. These form part of a wider set of arrangements for instructing, assessing, and counselling which embody core assumptions about how knowledge and pedagogy should be organized. This 'management' framework, in turn, is embedded within wider departmental and institutional structures, each with its own set of procedures, and professional and societal allegiances. Though apparently far removed from the assimilation and

schematization of knowledge at the classroom level, these 'higher-order' aspects of the school or college environment cannot be ignored. To take an example: teaching and learning in a particular setting are profoundly influenced by the type of assessment procedures in use; by constraints of scheduling; by the size and diversity of classes; by the availability of teaching assistants, library, computing, and copying facilities. These, in turn, are dependent on departmental priorities; on policies of faculty promotion; on institutional myths and traditions; and on local and national pressures.

The learning milieu concept is necessary for analysing the interdependence of learning and teaching, and for relating the organization and practices of instruction with the immediate and long-term responses of students. For instance, students' intellectual development cannot be understood in isolation but only within a particular school or college milieu. Equally, there are phenomena of crucial educational significance (such as boredom, interest, concentration, 'floundering', and intellectual dependency) that make nonsense of the traditional psychological distinction between 'cognitive' and 'affective', and which customarily arise as responses to the total learning milieu, not to single components of it. Students do not respond merely to presented content and to tasks assigned. Rather, they adapt to and work within the learning milieu taken as an inter-related whole. They pay close attention to 'hidden'[24] as well as

---

24. B.R. Snyder, The Hidden Curriculum (New York: Knopf, 1971).

'visible' curricula. Besides acquiring particular habits of studying, reading and responding, they also assimilate the conventions, beliefs, and models of reality that are constantly and inevitably transmitted through the total teaching process.[25]

ORGANIZATION AND METHODS OF ILLUMINATIVE EVALUATION

Illuminative evaluations - like the innovations and learning milieux that they study - come in diverse forms. The size, aims, and techniques of the evaluation depend on many factors: the sponsors' preoccupations; the exact nature and stage of the innovation; the number of institutions, teachers, and students involved; the level of cooperation and the degree of access to relevant information; the extent of the investigator's previous experience; the time available for data collection; the format of the required report; and, not least, the size of the evaluation budget.

Illuminative evaluation is not a standard methodological package but a general research strategy. It aims to be both adaptable and eclectic. The choice of research tactics follows not from research doctrine, but from decisions in each case as to the best available techniques: the problem defines the methods used, not vice versa.

---

25. For studies that examine various aspects of this 'secondary' learning and its relationship to intellectual development and social context see H.S. Becker et al., Making the Grade (New York: Wiley & Sons, 1968); W.G. Perry, Forms of Intellectual and Ethical Development in the College Years (New York: Holt, Rinehart & Winston, 1968); and M.F.D. Young op. cit..

Equally, no method (with its own built-in limitations) is used exclusively or in isolation; different techniques are combined to throw light on a common problem. Besides viewing the problem from a number of angles, this 'triangulation'[26] approach also facilitates the cross-checking of otherwise tentative findings.

At the outset, the researcher is concerned to familiarize himself thoroughly with the day-to-day reality of the setting or settings he is studying. In this he is similar to social anthropologists or to natural historians. Like them he makes no attempt to manipulate, control, or eliminate situational variables, but takes as given the complex scene he encounters. His chief task is to unravel it; isolate its significant features; delineate cycles of cause and effect; and comprehend relationships between beliefs and practices, and between organizational patterns and the responses of individuals. Since illuminative evaluation concentrates on examining the innovation as an integral part of the learning milieu, there is a definite emphasis both on observation at the classroom level and on interviewing participating instructors and students.

Characteristically in illuminative evaluation there are three stages: investigators observe, inquire further, and then seek to explain. Thus, in our study of a pilot project in independent learning

---

26. E.J. Webb et al., Unobtrusive Measures: Non-Reactive Research in the Social Sciences (Chicago: Rand McNally, 1966).

in British secondary schools,[27] early visits to the participating schools yielded a number of common incidents, recurring trends, and issues frequently raised in discussion. These were either observed ourselves, or heard about from teachers and pupils. (For example, we noticed that teachers spoke in different ways about the independent learning materials provided for use with their classes. While some regarded the sets of materials as constituting, collectively, a course of study, others saw the same materials as having a supplementary or ancillary function; to be used simply as a collection of resources to draw upon as, when, or if necessary.)

The second stage began with the selection of a number of such phenomena, occurrences, or groups of opinions as topics for more sustained and intensive inquiry. A change of emphasis accompanied this development. During the first, exploratory stage, we had become 'knowledgeable' about the scheme. At the second stage this enabled our questioning to be more focused; communication to be more coherent and relaxed; and, in general, observation and inquiry to be more directed, systematic, and selective. (Thus - in our contacts with the teachers - we sought to find out more about the status they assigned to the independent learning materials, and the extent to which they integrated them with others.)

---

27. The Nuffield Foundation Resources for Learning Project. The background to this is described in L.C. Taylor, Resources for Learning (Harmondsworth: Penguin Books, 1971). An outline of the associated evaluation research by M. Parlett will appear in L.C. Taylor (ed.), The Resources for Learning Handbook (in preparation).

The third stage consisted in seeking general principles underlying the organization of the program; spotting patterns of cause and effect within its operation; and placing individual findings within a broader explanatory context. It began with our weighing alternative interpretations in the light of information obtained. Thus, why did teachers differ in their attitudes towards the materials? It seemed in general that teachers' views depended on the availability of related materials in the school; on their previous experience with similar methods; and, - most critically - on whether or not they saw the material as 'displacing' or as 'supporting' the teacher. A number of other lines of investigation led to the same central issue: that of the changed role of the teacher in an independent learning setting.

Obviously the three stages overlap and functionally interrelate. The transition from stage to stage, as the investigation unfolds, occurs as problem areas become progressively clarified and re-defined. The course of the study cannot be charted in advance. Beginning with an extensive data base, the researchers systematically reduce the breadth of their enquiry to give more concentrated attention to the emerging issues. This 'progressive focusing' permits unique and unpredicted phenomena to be given due weight. It reduces the problem of data overload; and prevents the accumulation of a mass of unanalysed material.

Within this three-stage framework, an information profile is assembled using data collected from four areas: observation; interviews; questionnaires and tests; documentary and background sources.

Observation

As noted above, the observation phase occupies a central place in illuminative evaluation. The investigator builds up a continuous record of ongoing events, transactions and informal remarks.[28] At the same time he seeks to organize this data at source, adding interpretative comments on both manifest and latent features of the situation. In addition to observing and documenting day-to-day activities of the program, the investigator may also be present at a wide variety of other events (e.g. faculty and student meetings, open days, examiners' meetings, etc.).[29]

Much of the on-site observation involves recording discussions with and between participants. These provide additional information which might not otherwise be apparent or forthcoming from more formal interviews. The language conventions, slang, jargon, and metaphors that characterize conversation within each learning milieu, can reveal tacit assumptions, inter-personal relationships and status differentials.

Finally, there is a place for codified observation, using schedules for recording patterns of attendance, seating, utilization of

---

28. A useful source of participant observation research methods is G.J. McCall and J.L. Simmons, Issues in Participant Observation (London: Addison-Wesley, 1969).

29. For a research study that draws extensively on non-official, 'back-of-the-shop' settings see L.M. Smith & P.M. Keith, Anatomy of Educational Innovation (New York: John Wiley, 1971).

time and facilities, teacher-pupil interaction, etc.[30] The illuminative evaluator is cautious in the deployment of such techniques. In that they record only surface behavior they do not facilitate the uncovering of underlying, more meaningful features.

Interviews

Discovering the views of participants is crucial to assessing the impact of an innovation.[31] Instructors and students are asked about their work, what they think of it, how it compares with previous experiences; and also to comment on the use and value of the innovation. Interviews vary as to the type of information or comment that is sought. While brief, structured interviews are convenient for obtaining biographical, historical or factual information, more open-ended and discursive forms are suitable for less straightforward topics (e.g. career ambitions and anxieties).

Though desirable, it is rarely possible to inverview every participant, except in small innovatory programs or with large research teams. Interviewees, therefore, must usually be selected randomly or

---

30. For a discussion of classroom 'observation systems' see G.A. Nuthall, "A review of some selected studies of classroom interaction and teaching behaviour," in J.J. Gallagher et al., Classroom Observation, AERA Monograph No. 6, 1970.

31. Various approaches to interviewing can be found in the social sciences. Contrast the opposing perspectives presented by H.H. Hyman et al., (Interviewing in Social Research, Chicago: University of Chicago Press, 1954) and A.V. Cicourel (Method and Measurement in Sociology New York: Free Press, 1967). In that it is more characteristic of the 'anthropological' paradigm, illuminative evaluation favours the latter approach.

by 'theoretical' sampling.[32] This latter mode requires seeking out informants or particular groups who may have special insight or whose position makes their viewpoints noteworthy (e.g. students who have won prizes or failed altogether; marginal faculty members, who may have close knowledge of the innovation but have stayed outside it; young assistants teaching in their first semester, etc.). Those interviewed can also include more distant but equally relevant figures: e.g. at the college level, deans, administrators, and student counsellors; and, beyond the college, curriculum developers and foundation officials from whom the innovation stemmed.

Questionnaire and Test Data

While concentrating on observation and interviews, the illuminative evaluator does not eschew paper and pencil techniques. Their advantage in larger scale illuminative studies is especially evident. Also survey-type questionnaires used late in a study can sustain or qualify earlier tentative findings. Free and fixed response formats can be included to obtain both quantitative summary data and also open-ended (and perhaps new and unexpected) comment.[33]

There are, of course, several valid objections to questionnaires, particularly if they are used in isolation. Unless most carefully prepared, questionnaires can lead to mindless accumulations of uninterpretable

---

32. B. Glaser and A. Strauss, The Discovery of Grounded Theory (New York: Aldine, 1967).

33. If necessary, this qualitative data can be content analysed, to furnish further numerical results.

data. Expensive in time and resources, such careful preparation must be weighed against the benefits likely to accrue. A second drawback is that many recipients regard questionnaires as impersonal and intrusive. Others, keen to express their complicated views, find the questionnaire a frustrating, indeed trivializing medium. From these dissatisfied groups, some do not reply; yet these non-respondents may be the most important in certain respects.[34]

Besides completing questionnaires, participants can also be asked to prepare written comments on the program; to go through check-lists; or compile work diaries that record their activities over a specific period of time.[35]

Finally there are published or custom-built tests of attitude, personality, and achievement. Such tests enjoy no privileged status within the study. Test scores cannot be considered in isolation;[36] they form merely one section of the data profile. Interest lies not so much in relating different test scores, but in accounting for them using the study's findings as a whole.

---

34. In an unpublished questionnaire study at MIT, non-response was found to be the best predictor of student drop-out.

35. M. Parlett, "Classroom and Beyond: A Study of a Sophomore Physics Section at MIT," (MIT: Education Research Center, 1967).

36. '... Educators should continue to be apprehensive about evaluating teaching on the basis of performance testing alone. They should know how difficult it is to represent educational goals with statements of objectives. They should know how costly it is to provide suitable criterion testing. (And) They should know that the common-sense interpretation of these results is frequently wrong...' R.E. Stake, op. cit.

Documentary and Background Information

Innovations do not arise unheralded. They are preceded by committee minutes, funding proposals, architectural plans, and consultants' reports. Also other primary sources are obtainable: e.g. non-confidential data from registrars' offices; autobiographical and eye-witness accounts of the innovation; tape recordings of meetings; and examples of students' assignments.

The assembly of such information can serve a useful function. It can provide an historical perspective of how the innovation was regarded by different people before the evaluation began. The data may also indicate areas for inquiry (e.g. how representative were the students taking part?); may point to topics for intensive discussion (e.g. why were certain major features of the original proposal later abandoned?); or may expose aspects of the innovation that would otherwise be missed (e.g. why were subject requirements not fulfilled?).[37]

PROBLEMS AND POSSIBILITIES OF ILLUMINATIVE EVALUATION

First encounters with the radically different perspective of illuminative evaluation prompt a number of important questions.

---

37. These examples are drawn from an illuminative evaluation study of two innovative freshman programs at MIT (The Unified Science Study Program, and the Experimental Study Group). Each offered a full-time alternative to the traditional first year program. (M. Parlett, "Study of two experimental programs at MIT," unpublished, 1971).

a) Foremost is usually concern over the 'subjective' nature of the approach. Can 'personal interpretation' be scientific? Is not collection, analysis, and reporting of data, sceptics ask, entirely at the discretion of the researchers themselves?

Behind such questions lies a basic but erroneous assumption: that forms of research exist which are immune to prejudice, experimenter bias, and human error. This is not so. Any research study requires skilled human judgements and is thus vulnerable.[38] Even in evaluation studies that handle automatically-processed numerical data, judgement is necessary at every stage: in the choice of samples; in the construction or selection of tests; in deciding conditions of administration; in selecting the mode of statistical treatment (e.g. whether or not to use factor analysis); in the relative weight given to different results; and, particularly, in the selection and presentation of findings in reports.

Nevertheless, the extensive use of open-ended techniques, progressive focusing, and qualitative data in illuminative evaluation still raises the possibility of gross partiality on the part of the investigator. A number of precautionary tactics are possible. During the investigation different techniques can be used to cross-check the most important findings; open-ended material can be coded and checked by outside researchers; consultants to the evaluation can be charged

---

38. For a general discussion of this problem area see R. Rosenthal, Experimenter Effects in Behavioral Research (New York: Appleton-Century-Crofts, 1966).

with challenging preliminary interpretations and playing devil's advocate; and members of the research team can be commissioned to develop their own interpretations.[39] At the report stage, in addition to the findings, critical research processes can also be documented: theoretical principles and methodological ground rules can be discussed and made explicit; criteria for selecting or rejecting areas of investigation can be spelled out; and evidence can be presented in such a way that others can judge its quality.

Even with such precautions, the subjective element remains. It is inevitable. When the investigator abandons the agricultural-botany paradigm his role is necessarily re-defined. The use of interpretative human insight and skills is, indeed, encouraged rather than discouraged. The illuminative evaluator thus joins a diverse group of specialists (e.g. psychiatrists, social anthropologists and historians), where this is taken for granted. In each of these fields the research worker has to weigh and sift a complex array of human evidence and draw conclusions from it.

A further issue also focuses on the position of the investigator. Does not his presence have an effect on the conduct and progress of the innovatory scheme he is studying? Certainly it does; indeed, any form of data collection creates disturbance. Illuminative evaluators recognize this and attempt to be unobtrusive without being secretive;

---

39. The added possibility of research 'in tandem' with different investigators working in semi-isolation and pooling their findings at the end, is currently being examined with respect to a proposed British evaluation involving the authors.

to be supportive without being collusive; and to be non-doctrinaire without appearing unsympathetic.

This leads to an important point: that research workers in this area need not only technical and intellectual capability, but also inter-personal skills. They seek cooperation but cannot demand it. There may be times when they encounter nervousness and even hostility. They are likely to be observing certain individuals at critical times in their lives (e.g. students about to leave, or instructors with a high personal investment in the innovation). The researchers need tact and a sense of responsibility similar to that pertaining in the medical profession. They seek and are given private opinions, often in confidence. They are likely to hear, in the course of their study, a great deal about personalities and institutional politics that others might be inquisitive to know. There are especially difficult decisions to make at the report stage: though full reporting is necessary, it is essential to safeguard individuals' privacy.

Such problems, confronting many research workers in the human sciences, are exacerbated in the case of close-up, intensive studies of the type outlined here. The price of achieving the richer, more informative data of illuminative evaluation is the greatly increased attention that must be paid to the evaluator's professional standards and behavior. Though there can be no fixed rules, there are certain guidelines for the illuminative evaluator. For instance, to retain the viability and integrity of his research position and the trust of the participants in the program, the investigator needs, from the outset, to clarify his role; to be open about the aims of his study; and to

ensure that there is no misunderstanding or ambiguity about who, for example, will receive the report.[40]

b) Besides concern with the investigator's special position, illuminative evaluation also prompts questions concerning the scope of the investigation. Is illuminative evaluation confined to small scale innovations. Can it be applied to innovations that are being widely implemented? Detailed studies of specific learning milieux may be insightful and valid, but are the results and analyses generalizable to other situations? Is it possible to move from the particular to the universal?

Despite its basis in the close-up study of individual learning milieux, illuminative evaluation can also be applied on a wider scale. Suppose an innovatory program had been adopted by many different schools. At the beginning of the evaluation a small sample of schools could be selected for intensive study. As the study progressed, and as it focused on selected salient issues arising in the different learning milieux, the number of schools studied could be expanded. The new investigations, now more selective, could be pursued more speedily, with concentration more on noting similarities and differences between

---

40. He must also not be a 'snooper', nor become an 'institutionalized voyeur' in succumbing to his private research interests. He should also avoid the dangers of 'going native' or getting caught up in political intrigues. For an early but still relevant discussion of these problems, see various papers in R.N. Adams & J.J. Preiss (eds.), Human Organization Research (Homewood, Illinois: Dorsey Press, 1960).

situations, than on full documentation of each learning milieu.[41] Finally, with this further information assimilated, short visits, or even - in the last resort - mailed questionnaires could be used for the remainder of the institutions.

The full progression - from small sample studies to larger scale inquiries - is often only necessary in widely applied programs. But there is another way in which perceptive and rigorous study of specific situations can yield more generally applicable insights with either large or small-scale investigations. Learning milieux, despite their diversity, share many characteristics. Instruction is constrained by similar conventions, subject divisions, and degrees of student involvement. Teachers encounter parallel sets of problems. Students' learning, participation, study habits, and examination techniques are found to follow common lines; and innovations, as such, face habitual difficulties and provoke familiar reactions. There is a wide range of overlapping social and behavioral phenomena that accompany teaching, learning, and innovating. This is widely acknowledged. However, few of these phenomena have been pinpointed, adequately described or defined accurately. Illuminative evaluation aims to contribute to this process. There is a need for abstracted summaries, for shared terminology, and for insightful concepts. These can serve as aids to communication and facilitate theory-building. They have been conspicuously

---

41. At the same time it is necessary to remain extremely flexible and to be open to new issues that arise in the later stages of a study.

absent from most research in education. Yet, without this conceptual equipment, the universals of teaching will be cyclically discovered, described, forgotten, re-discovered, and described again.

DECISION-MAKING, EVALUATION, AND ILLUMINATION

The principal purpose of evaluation studies is to contribute to decision-making.[42] There are at least three separate but related groups of decision-makers to whom the evaluator addresses his report: (i) the program's participants; (ii) the program's sponsors, supervisory committee, or educational board; (iii) interested outsiders (such as other researchers, curriculum planners, etc.).

Each group or constituency will look to the report for help in making different decisions. The participants, for example, will be anxious to correct deficiencies, make improvements and establish future priorities. The sponsors and board members will be concerned with pedagogic issues but will also want to know about the innovation's costs, use of resources, and outside reputation. The outsiders will read the report to decide whether or not the scheme has 'worked', or to see whether it could be applied or adapted to their own situations.

Clearly, if the evaluator is to acknowledge the interests of all

---

42. In practice, motives for commissioning evaluations are often mixed. Some 'evaluations' may be used to delay troublesome decisions or to window-dress a policy already formulated. Exceptionally they may be instigated simply to satisfy a funding agency's demands.

these groups, he cannot - even if requested - provide a simple 'yes' or 'no' on the innovation's future. A decision based on one group's evaluative criteria would, almost certainly, be disputed by other groups with different priorities. A 'mastery of fundamentals' for one group, is for another a 'stifling of creativity'. The investigator does not make decisions. Indeed, in these terms he cannot - except as a representative or agent of one of the interest groups.[43]

Illuminative evaluation thus concentrates on the information-gathering rather than the decision-making component of evaluation. The task is to provide a comprehensive understanding of the complex reality (or realities) surrounding the program: in short, to 'illuminate'. In his report, therefore, the evaluator aims to sharpen discussion, disentangle complexities, isolate the significant from the trivial, and to raise the level of sophistication of debate.

SUMMARY

When an innovation ceases to be an abstract concept or plan, and becomes part of the teaching and learning in a school or college, it assumes a different form altogether. The theater provides an analogy: to know whether a play 'works' one has to look not only at the manuscript but also at the performance; that is, at the

---

43. If the evaluator allows his study to be defined in this way - oriented towards one constituency only - he accepts the more limited role and quite different tasks of the 'service' researcher.

32

interpretation of the play by the director and actors. It is this that is registered by the audience and appraised by the critics. Similarly, it is not an instructional system as such but its translation and enactment by teachers and students, that is of concern to the evaluator and other interested parties. There is no play that is 'director-proof'. Equally, there is no innovation that is 'teacher-proof' or 'student-proof'.

If this is acknowledged, it becomes imperative to study an innovation through the medium of its performance and to adopt a research style and methodology that is appropriate.

This involves the investigator leaving his office and computer print-out to spend substantial periods in the field. The crucial figures in the working of an innovation - learners and teachers - become his chief preoccupation. The evaluator concentrates on 'process' within the learning milieu, rather than on 'outcomes' derived from a specification of the instructional system.[44] Observation, linked with discussion and background inquiry enable him to develop an informed account of the innovation in operation.

Ideally, the output of his research will be regarded as useful, intelligible and revealing, by those involved in the enterprise itself. Further, by addressing key educational issues it can also be seen as a recognizeable reality by others outside the innovation. If the report

---

44. An agricultural-botany evaluator is rather like a critic who reviews a production on the basis of the script and applause-meter readings, having missed the performance.

is seen merely as an arcane or irrelevant addition to a research literature already ignored by practising educators, clearly the evaluator will have failed.

In attempting to document the teacher-student interactions, intellectual habits, institutional constraints, etc., that characterize classroom life, the investigator contributes to a field that has received only minimal attention from social scientists.[45] Until recently, perceptive accounts of learning milieux have, more often than not, been found in 'travellers' tales'[46] or 'non-fiction' novels[47] rather than in educational research reports. The investigator has, therefore, not only short-term goals, but also the long-term goal of contributing to a developing and urgently required new field of study.

This approach does not cure all ills, nor can any one approach. Certainly, no simplified instant solutions to perennial educational questions will be delivered by such studies. Indeed, by discarding a spurious 'technological' simplification of reality, and by acknowledging the complexity of educational process, the illuminative evaluator is likely to increase rather than lessen the sense of uncertainty in

---

45. This paper has focused on the evaluation of innovatory programs. There is an obvious need (not always acknowledged) for comparable studies to be made of traditional teaching. Illuminative evaluation need not be confined to innovation.

46. e.g., J. Holt, How Children Fail (New York: Dell, 1964).

47. e.g., J. Herndon, The Way It Spozed To Be (New York: Simon & Schuster, 1965).

education. On the other hand, unless studies such as these are vigorously pursued there is little hope of ever moving beyond helpless indecision or doctrinal assertion in the conduct of instructional affairs.

The authors are grateful to the following for their comments on a previous draft: Clem Adelman, Tony Becher, Sara Delamont, Oliver Fulton, Liam Hudson, Carolyn Miller, Peter Sheldrake, Deirdre Snyder, Kim Taylor, Brian Torode, Martin Trow, Rob Walker and members of M. Parlett's seminar at the 1972 Danforth Foundation Workshop on Liberal Arts Education at Colorado Springs.

The preparation of this paper has been supported by a grant from the Nuffield Foundation, London.